Research Paper | Computer Science | E-ISSN No : 2454-9916 | Volume : 10 | Issue : 3 | March 2024

ABSTRACTIVE SUMMARIZATION OF INDIAN LEGAL DOCUMENTS USING T5 & QLORA

Dr. M. S. Anbarasi[1], Aathif Mohammed A.[2], Deena R.[3], Manimaran V.[4], Mohanraj S.[5]

[1, 2, 3, 4, 5] Puducherry Technological University

ABSTRACT

This research project aims to develop an abstractive summarization system for Indian legal documents. The system leverages the power of the T5 transformer model, fine-tuned using Quantized Low-Rank Adaptation (QLoRA). The training data comprises two datasets, the Indian Legal Corpus (ILC) and IN-Abs, both containing court cases and their corresponding abstractive summaries.

The system is designed to accept legal text input directly or extract it from uploaded DOCX or PDF documents. An initial extractive summary is generated using the bert-extractive-summarizer, which is subsequently fed into the fine-tuned T5 model to produce an abstractive summary.

The principal result of this research is the successful implementation of a system capable of generating abstractive summaries of Indian legal documents. The system achieved a ROUGE-1 score of 46.37%, demonstrating its effectiveness.

In conclusion, this research contributes to the field of legal document summarization by providing a system that can generate concise and coherent summaries, thereby aiding in the efficient comprehension of complex legal texts. This work also opens avenues for further improvements and applications in the legal tech domain.

KEYWORDS: Abstractive Summarization, Indian Legal Documents, T5 Transformer Model, Quantized Low-Rank Adaptation (QLoRA), Bert-Extractive-Summarizer, ROUGE-1 Score

INTRODUCTION

The field of legal document summarization has seen significant advancements with the advent of transformer-based models. However, the complexity and specificity of Indian legal documents present unique challenges that necessitate specialized solutions. This study introduces an innovative approach to abstractive summarization of Indian legal documents, building upon recent advancements in transformer models and fine-tuning techniques.

The purpose of this research is to develop a system capable of generating abstractive summaries from Indian legal documents. The system employs the T5 transformer model, which has shown promising results in various natural language processing tasks. To adapt the model to the specific task and data, we use Quantized Low-Rank Adaptation (QLoRA), a fine-tuning technique that has demonstrated effectiveness in similar applications.

The T5 model is trained on two datasets, the Indian Legal Corpus (ILC) and IN-Abs, both of which contain court cases and their abstractive summaries. The use of these datasets ensures that the model is well-suited to handle the intricacies of Indian legal texts.

The system accepts legal text directly or extracts it from uploaded DOCX or PDF documents. An initial extractive summary is generated using the bert-extractive-summarizer, which is then fed into the fine-tuned T5 model to produce an abstractive summary.

The results of this research indicate that the system can effectively generate abstractive summaries, achieving a ROUGE-1 score of 46.37%. This study contributes to the ongoing efforts in the field of legal document summarization and opens up new possibilities for future research and applications in the legal tech domain.

MATERIALS AND METHODS

T5, short for Text-to-Text Transfer Transformer, is a versatile neural network model developed by Google for various natural language processing tasks.[2] It operates on a text-to-text approach, converting both input and output into text[3]. This makes T5 flexible for tasks like translation, summarization, sentiment classification, and more. T5 uses an abstractive summarizing algorithm, generating new sentences from the given text. It requires the text to be transformed into numerical form for training and inference. This powerful model has significantly impacted the field of NLP, offering a unified framework for diverse tasks.

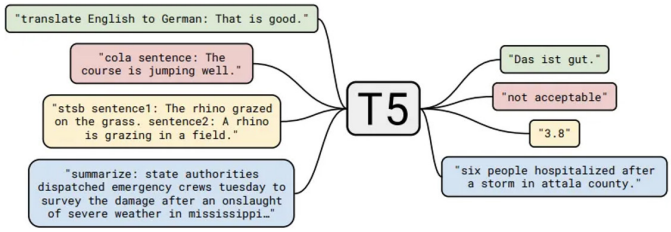

Figure 1: T5's text-to-text framework

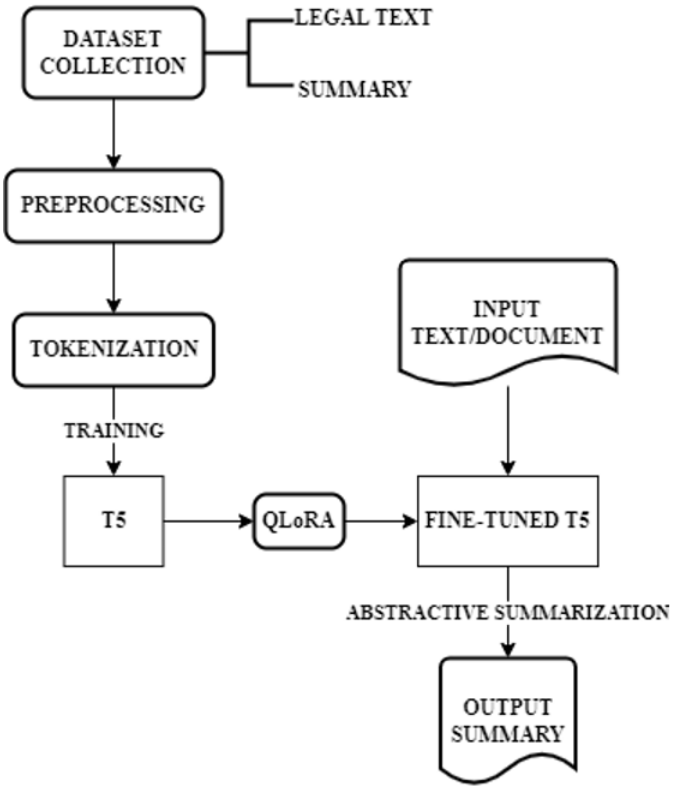

Figure 2: Proposed system

This project employs the T5 transformer model for the abstractive summarization of Indian legal documents. The model is fine-tuned using Quantized Low-Rank Adaptation (QLoRA), a method that adapts the pre-trained model to the specific task of summarizing legal documents.

Datasets

The model was trained on two datasets: the Indian Legal Corpus (ILC) and IN-Abs. Both datasets contain Indian court cases along with their abstractive summaries.

Document Processing

Users can input the legal text directly or upload it as a document in DOCX or PDF format. The text is then extracted from these documents and sent to the BERT Extractive Summarizer. [9] This summarizer generates an extractive summary, which serves as the input for our fine-tuned model.

Model Fine-tuning and Summarization

Quantized Low-Rank Adaptation (QLoRA) is an efficient fine-tuning approach that significantly reduces memory usage, enabling the fine-tuning of large models on a single GPU.[3] It backpropagates gradients through a frozen, 4-bit quantized pretrained language model into Low Rank Adapters (LoRA). QLoRA introduces several innovations to save memory without sacrificing performance, such as 4-bit NormalFloat (NF4), a new data type optimal for normally distributed weights, and double quantization to reduce the average memory footprint. It has been used to fine-tune more than 1,000 models, achieving state-of-the-art results.

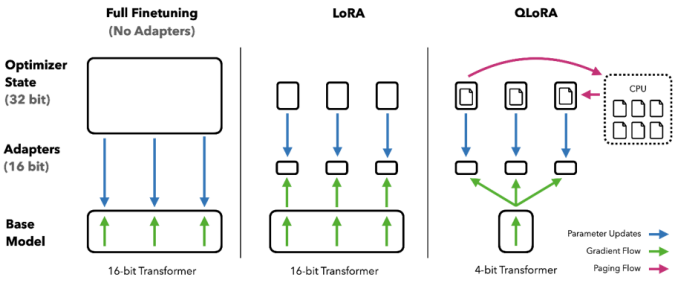

Figure 3: Different finetuning methods and their memory requirements. QLORA improves over LoRA by quantizing the transformer model to 4-bit precision and using paged optimizers to handle memory spikes.

The extractive summary is fed into the fine-tuned T5 model, which generates the final abstractive summary. The fine-tuning process involves training the T5 model on our datasets using QLoRA, which adapts the model to the specific task of summarizing legal documents.

Evaluation

The performance of the model was evaluated using the ROUGE-1 score, a common metric for evaluating summarization models. Our model achieved a ROUGE-1 score of 46.37%, indicating a high level of accuracy in generating abstractive summaries.

RESULTS AND DISCUSSION

The project was designed with the aim of creating an abstractive summarization model for Indian legal documents. The T5 transformer model was chosen for this task due to its proven effectiveness in text summarization tasks. The model was fine-tuned using Quantized Low-Rank Adaptation (QLoRA), a method that adapts the pre-trained model to the specific task of summarizing legal documents.

The model was trained on two datasets: the Indian Legal Corpus (ILC) and IN-Abs. Both datasets contain Indian court cases along with their abstractive summaries. The training process involved optimizing the model parameters to minimize the loss function, which measures the difference between the model's predictions and the actual summaries. The fine-tuned model achieved a training loss of 1.98.

MODEL	ROUGE-1
T5 base on ILC & IN-Abs	8.01%
Fine-tuned T5 on ILC & IN-Abs	**46.37%**

Table 1: Results

The validation of the model was performed by testing it on unseen legal documents and comparing the generated summaries with the actual summaries. The model achieved a ROUGE-1 score of 46.37%, a significant improvement over the base T5 model, which achieved a ROUGE-1 score of 8.01% on the test set.

The results demonstrate the effectiveness of the T5 model in summarizing Indian legal documents when fine-tuned using QLoRA. The high ROUGE-1 score indicates that the model was able to generate summaries that closely match the actual summaries.

The use of the BERT Extractive Summarizer to generate an extractive summary, which serves as the input for our fine-tuned model, proved to be an effective strategy. This approach allowed the model to focus on the most important parts of the document, thereby improving the quality of the abstractive summary.

However, it's important to note that while the model achieved a high ROUGE-1 score, there is still room for improvement. Future work could explore other fine-tuning methods or use additional datasets to further improve the model's performance.

CONCLUSION

This research project successfully developed an abstractive summarization model for Indian legal documents using the T5 transformer model, fine-tuned with Quantized Low-Rank Adaptation (QLoRA). The model was trained on two datasets, the Indian Legal Corpus (ILC) and IN-Abs, both containing Indian court cases and their abstractive summaries.

The unique approach of using the BERT Extractive Summarizer to generate an extractive summary, which was then used as input for the fine-tuned T5 model, proved to be effective. This strategy allowed the model to focus on the most important parts of the document, thereby enhancing the quality of the abstractive summary.

The model achieved a ROUGE-1 score of 46.37%, demonstrating its effectiveness in generating summaries that closely match the actual summaries. This is a significant improvement over the base T5 model, which achieved a ROUGE-1 score of 8.01% on the test set.

While the results are promising, there is still room for improvement. Future work could explore other fine-tuning methods or use additional datasets to further enhance the model's performance. This project lays a solid foundation for future research in the field of legal document summarization. It has the potential to significantly contribute to the efficiency and accessibility of legal proceedings, thereby having a profound impact on the legal system.

REFERENCES

1. Ay, B., Ertam, F., Fidan, G., & Aydin, G. (2023). Turkish abstractive text document summarization using text to text transfer transformer. Alexandria Engineering Journal, 68, 1-13.
2. A'yuna Itsnaini, Q., Hayaty, M., Putra, A. D., & Jabari, N. A.M. (2023). Abstractive Text Summarization using Pre-Trained Language Model "Text-to-Text Transfer Transformer (T5)". ILKOM Jurnal Ilmiah, 15(1), 124-131.
3. Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). QLORA: Efficient Finetuning of Quantized LLMs. In 37th Conference on Neural Information Processing Systems (NeurIPS 2023).
4. Ghosh, S., Dutta, M., & Das, T. (2022). Indian Legal Text Summarization: A Text Normalisation-based Approach. Paper presented at the IEEE 19th India Council International Conference (INDICON). doi:10.36227/techrxiv.19944665.v4
5. Itsnaini, Q. A'yuna, Hayaty, Mardhiya, Putra, Andriyan Dwi, Jabari, Nidal A.M., Trivedi, Pawan, Jain, Digha, Gite, Shilpa, Kotecha, Ketan, Bhatt, Anant, & Naik, Nithesh. (2023). Indian Legal Corpus (ILC): A Dataset for Summarizing Indian Legal Proceedings using Natural Language. Engineered Science. DOI: https://dx.doi.org/10.30919/es1022
6. Mullick, A., Nandy, A., Kapadnis, M. N., Patnaik, S., Raghav, R., & Kar, R. (2022). An Evaluation Framework for Legal Document Summarization. In Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022) (pp. 4747–4753). Marseille, 20-25 June 2022. European Language Resources Association (ELRA).
7. Shruthi SK, Shwetha SK, & Ramya B. (2023). T5-Precis: Concise and Precise Text Summarization. International Journal of Research Publication and Reviews, 4(12), 1235-1240. ISSN 2582-7421.
8. SKT5SciSumm - A Hybrid Generative Approach for Multi-Document Scientific Summarization. (2024). Retrieved from arXiv: 2402.17311v1 [cs.CL].
9. Sophie, S. L. M., & Sathya, S. S. (2022). Extractive - Abstractive Summarization Using Transformers: A Hybrid Approach. Journal of Pharmaceutical Negative Results, 13(Special Issue 10), Volume 13.
10. Suryawanshi, V., Naikwadi, D., & Patil, S. (2023). Legal case document summarization using NLP. International Research Journal of Modernization in Engineering Technology and Science, 5(12), 1-3.
11. To, H. Q., Tran, H.-N., Greiner-Petter, A., Beierle, F., & Aizawa, A. (2024). SKT5SciSumm - A Hybrid Generative Approach for Multi-Document Scientific Summarization. arXiv:2402.17311v1 [cs.CL].
12. Tran, N., Prijs, D., Schraagen, M., & Bex, F. (2022). Abstractive Summarization of Dutch Court Verdicts Using Sequence-to-sequence Models. In Proceedings of the Natural Legal Language Processing Workshop 2022 (pp. 76-87). December 8, 2022. ©2022 Association for Computational Linguistics.
13. van de Luijtgaarden, N., Prijs, D., Schraagen, M., & Bex, F. (2022). Abstractive Summarization of Dutch Court Verdicts Using Sequence-to-sequence Models. In Proceedings of the Natural Legal Language Processing Workshop 2022 (pp. 76-87). December 8, 2022. ©2022 Association for Computational Linguistics.
14. Wang, M., Xie, P., Du, Y., & Hu, X. (2023). T5-Based Model for Abstractive Summarization: A Semi-Supervised Learning Approach with Consistency Loss Functions. Appl. Sci., 13, 7111. doi: 10.3390/app13127111.